**Article:**
Zhang, Zhihong, Liang, Yuanheng, Bai, Lu et al. (1 more author) (2016) Discriminative sparse representation for face recognition. Multimedia Tools and Applications. 3973–3992. ISSN: 1573-7721

https://doi.org/10.1007/s11042-015-3136-x

**Noname manuscript No.**
(will be inserted by the editor)

# Discriminative Sparse Representation for Face Recognition

**Zhihong Zhang** [1] · **Yuanheng Liang** [2] · **Lu Bai** [3] · **Edwin R. Hancock** [4]

Received: date / Accepted: date

**Abstract** Recently Sparse Representation (or coding) based Classification (SRC) has gained great success in face recognition. In SRC, the testing image is expected to be best represented as a sparse linear combination of training images from the same class, and the representation fidelity is measured by the $\ell_2$-norm or $\ell_1$-norm of the coding residual. However, SRC emphasizes the sparsity too much and overlooks the spatial information during local feature encoding process which has been demonstrated to be critical in real-world face recognition problems. Besides, some work considers the spatial information but overlooks the different discriminative ability in different face regions. In this paper, we propose to weight spatial locations based on their discriminative abilities in sparse coding for robust face recognition. Specifically, we learn the weights at face locations according to the information entropy in each face region, so as to highlight locations in face images that are important for classification. Furthermore, in order to construct a robust weights to fully exploit structure information of each face region, we employed external data to learn the weights, which can cover all possible face image variants of different persons, so the robustness of obtained weights can be guaranteed. Finally, we consider the group structure of training images (i.e. those from the same subject) and added an $\ell_{2,1}$-norm (group Lasso) constraint upon the formulation, which enforcing the sparsity at the group level. Extensive experiments on three benchmark face datasets demonstrate that our proposed method is much more robust and effective than baseline methods in dealing with face occlusion, corruption, lighting and expression changes, etc.

**Keywords** Sparse coding · Coding fidelity · Face recognition · Spatial information

Corresponding Author: Yuanheng Liang, Email: 990042350@qq.com

[1] Software school, Xiamen University, Xiamen, Fujian, China
[2] School of Mathematical Sciences, Xiamen University, Xiamen, Fujian, China
[3] School of Information,Central University of Finance and Economics, Beijing, China
[4] Department of Computer Science, University of York, York, UK

## 1 Introduction

Face recognition has received continuous attentions for several decades as it is an important topic in computer vision. Numerous methods have been proposed to transform face image data to a lower dimensional feature space for recognition, such as Eigenfaces [2], Fisherfaces [2], locality preserving projection (LPP) [8] and Laplacianfaces [9]. Moreover, to deal with practical face recognition problems, LBP [1] and its variants were used to deal with illumination changes. Although much progress have been made, robust face recognition remains a very challenging task, since real-world face images contain noisy background cluster and are typically with significant lighting, expression, pose, etc. variations.

Representation based face recognition methods have been recently proposed for robust face recognition [19, 6, 21, 13]. One typical example is the sparse representation based representation (SRC) scheme [19], which converts the query face image into a sparse linear combinations of training images with illumination, etc. variations. By imposing an $\ell_1$-norm constraint on the resulting coefficients, SRC achieved very promising results on face recognition, but it needs much computational cost. Zhang et al. [24] then proposed to use the $\ell_2$-norm to regularize the representation coefficients which achieves similar accuracy to SRC but with significantly less computational cost.

One major computational problem of sparse coding is to improve the quality of the sparse representation while maximally preserving the signal fidelity. To achieve this goal, many works have been proposed to modify the sparsity constraint. For example, Yang et al. [20] proposed to use robust sparse coding along with max pooling for image classification and achieved good performance over traditional $k$-means clustering based method. Liu et al. [11] added nonnegative constrain to the sparse coefficients. Wang et al. [17] used locality constraints during the sparse coding process to speed up computation and coding efficiency. To maintain similarity, Gao et al. [5] introduced a Laplacian term of coefficients in sparse coding, which was extended to an efficient algorithm in Cai et al. [25]. In addition, Ramirez et al. [15] proposed a framework of universal sparse modeling to design sparsity regularization terms. The Bayesian methods were also used for designing the sparsity regularization terms [10].

The above developments of sparsity regularization term improve the sparse representation in different aspects; However, to the best of our knowledge, little work has been done on improving the fidelity term except those in [18, 19, 21, 22]. In [18, 19], the $\ell_1$-norm was used to define the coding fidelity. In [21, 22], they design the signal fidelity term as an maximum likelihood (MLE) estimation or maximum a posterior (MAP) estimation, which minimizes some function (associated with the distribution of the coding residuals) of the coding residuals. In fact, the fidelity term has a high impact on the final coding results because it ensures that the given signal $y$ can be faithfully represented by the dictionary $D$. Although the effectiveness of these methods have been proven, the spatial information is lost during the coding phase. We believe the spatial information should also be included in the coding process. Some works have demonstrate the importance of spatial information [23, 17]. However, they ignores the fact that different spatial regions in a face image may have distinct discriminative abilities.

To improve the robustness and effectiveness of sparse representation, we propose to incorporated the discriminative ability of pixel locations into the sparse coding procedure. We believe that the amount of information in different face regions is different. By intuition, some regions (such as mouth, eyes, nose) rich in texture should contain more information; thus, they are expected to be assigned with high weight values to ensure very small residuals while others, like cheek, can be almost homogenous, are given lower weight values to reduce their effects on the regression estimation so that sensitiveness to these regions can be greatly reduced. Such weight values are determined through the information entropy in each face region. In order to construct a robust weights to fully exploit structure information of each face region, we employed external data (not just limit to training data) to learn the weights. As the external data can cover all possible face image variants of different persons, so the robustness of obtained weights can be guarantee.

## 2 Brief Review of Sparse Coding Model

The goal of sparse coding is to learn a dictionary and corresponding sparse codes such that input data can be well approximated [14]. The traditional sparse coding model can be interpreted in the following optimization problem:

$$\min_{\boldsymbol{\alpha}} \|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{\alpha}\|_2^2 \quad s.t. \quad \|\boldsymbol{\alpha}\|_1 \leq \sigma \tag{1}$$

$$\min_{\boldsymbol{\alpha}} \|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{\alpha}\|_1 \quad s.t. \quad \|\boldsymbol{\alpha}\|_1 \leq \sigma \tag{2}$$

where $\sigma > 0$ is a constant, $\boldsymbol{y} = [y_1; y_2; \cdots; y_n] \in \Re^n$ is the signal to be coded, $\boldsymbol{D} = [d_1, d_2, \cdots, d_m] \in \Re^{n \times m}$ represents the dictionary, which can be either learned or predefined, and $\alpha$ is the estimated coefficient which is supposed to be sparse. In Eq.1, the coding residual is modeled by $\ell_2$-norm, which is a formulation of so-called LASSO problem. While Eq.2 uses $\ell_1$-norm to model the representation residual for robustness to occlusion, this will increase much the computational cost. From the viewpoint of maximal likelihood estimation, the $\ell_2$-norm and $\ell_1$-norm characterization of the representation residual is only optimal when the coding residual $\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{D}\boldsymbol{\alpha}$ follows Gaussian or Laplacian distribution.

In practice, however, both the Gaussian and Laplacian assumptions of the distribution of the residual $e$ may not be appropriated when the face images are subject to complex variations, such as occlusions, corruptions, or expression variations. To solve this problem, Yang et al. [21] proposed robust sparse coding (RSC) to measure the representation residual. In RSC, the reconstruction error is $e_i = y_i - d_i\alpha_i$, $i = 1, 2, \ldots, n$ and $\{e_1, e_2, \ldots, e_n\}$ are assumed to be independent with some probability density function not necessarily Gaussian or Laplacian. The maximum likelihood estimation principle is utilized to robustly represent the given signal with sparse regression coefficients. To the end, they transformed the optimization problem into an iteratively reweighted sparse coding problem. The sparse coding model can be written as:

$$\min_{\boldsymbol{\alpha}} \|\boldsymbol{M}^{\frac{1}{2}}(\boldsymbol{y} - \boldsymbol{D}\boldsymbol{\alpha})\|_2^2 \quad s.t. \quad \|\boldsymbol{\alpha}\|_1 \leq \sigma \tag{3}$$

where $\boldsymbol{M}$ is a diagonal matrix of weights assigned to pixels of the query image $\boldsymbol{y}$. For example, the elements $M_{i,i}$, i.e., $m_\theta(e_i)$, is the weight assigned to pixel $i$ of query image $y$. Yang et al. [21] choose to use the logistic function as the weight function

$$m_\theta(e_i) = \exp(\mu\eth - \mu e_i^2)/(1 + \exp(\mu\eth - \mu e_i^2)) \tag{4}$$

where $\mu$ and $\eth$ are positive scalars. For pixels corresponding to outliers (such as occlusion, corruption) and therefore with large residuals, the related elements in $\boldsymbol{M}$ will be adaptively suppressed to reduce their impacts on the regression estimation and improve the robustness to outliers.

Although the effectiveness of RSC has been proven, the spatial information is lost during the coding phase. We believe that the amount of information in different face regions is different and the spatial information should also be included in the coding process.

## 3 Information Entropy based Spatial Weighting Strategy

Note that in RSC approach (see Eq.3), the sparse coding is actually weighted pixel-wise. Elements at the diagonal of $\boldsymbol{M}$ correspond to weights resulting form outlier detection, which tend to have small values for pixels corresponding to outliers. However, RSC ignores the fact that different face regions may have different representation accuracy and contributions to face recognition. By intuition, some regions rich in texture, like areas around eyes, are expected to have higher representation accuracies than the homogenous regions, like cheek. Therefore, if we can detect the important regions in face images and assign them with high weight values, more robust face recognition results can be obtained.

In order to make better use of the spatial information, traditional methods firstly divide the face images into rectangular regions. However, it is very difficult to decide the number and size of regions, especially when there are different appearance variations on the face. A finer division usually makes the descriptor more discriminative but sometimes, for example when there are expression variations, will bring some problems. This is because in the case of expression variations, small regions around some face areas, such as mouth and eyes, are shifted to neighbor regions.

Inspired from the recent development on regions division problem [3], we learn the weight for each region according to the information entropy of the region. More specifically, we firstly divided the face images into a few coarse rectangular regions and then the pixels in each region are regrouped into different sets by information entropy measure. This strategy allows to group most of the relevant pixels into corresponding sets even in the presence of some shifting. Moreover, the importance for different pixel sets are their information entropy values. Figure.1 shows the details for learning the weight values in each face block using information entropy.

The entropy is a term defined in information theory as a measurement of the uncertainty associated with a random variable[4]. It is relevant to the quantity and variability of the information. Here, we assume that the pixel intensity value is a random variable; thus, we can use the histogram of intensities in each face region to approximate the probability density function (PDF) for computing the information

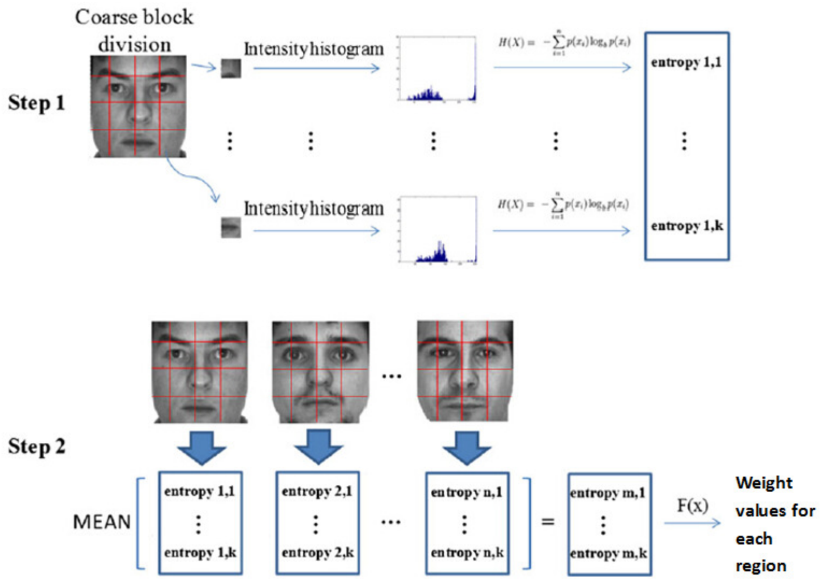


**Fig. 1** The details for learning the weight values in each face block using information entropy.

entropy. Applied to our case, the larger the entropy values is, the more information a face region should contain, and thus more clusters should be set for this region.

The entropy value of the face region $i$ can be then defined as

$$H_i = \sum_{k=1}^{n} p(v_k)\log_2\left(\frac{1}{p(v_k)}\right) = -\sum_{k=1}^{n} p(v_k)\log_2 p(v_k) \tag{5}$$

where $p(v_k)$ is the probability of the pixel $v$ with intensity value $k$ in the histogram of the region.

In our method, the entropy following Eq.5, is computed from the intensity histograms of the coarse divided regions for all face images in the training set. Then, the average entropy value of a region in all images is used as the corresponding regional entropy. Although some images in the training set might be affected by noise, the average entropy values can still reflect the information quantity differences among different facial regions. Finally, a monotonic transform function is used for mapping the entropy value to the final weight values.

The monotonic transform function $F(\hat{H}_i)$ in this paper, is implemented by using a linear function as follows:

$$N(\hat{H}_i) = (\hat{H}_i - H^{min})/(H^{max} - H^{min}) \times (new_{max} - new_{min}) + new_{min} \tag{6}$$

where $\hat{H}_i$ is the average entropy for region $i$, $H^{min}$ and $H^{max}$ are the minimum and the maximum entropy values from all regions, $new_{min}$ is the least weights the region should be set while $new_{max}$ is the maximum weight can be obtained in a region. If the output $N(\hat{H}_i)$ is not an integer number, it can be rounded to be an integer value. In this work, we have decided to use $new_{min} = 0.7$ and $new_{max} = 1.3$ and a coarse regions division of $6 \times 6$, aiming at having a good trade-off between the computation cost and the proper use of the local spatial information.

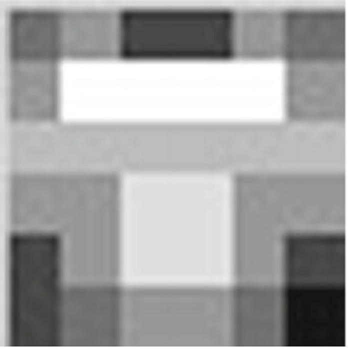


**Fig. 2** The learning results of set number for each face region. The set number is then corresponding to the weight for each face region. The brightest part represent those regions with high weight while the darkest parts correspond to the low weight.

Once the number of pixel sets $N(\hat{H}_i)$ in each region is learned by Eq.6, all the pixels on region $i$ are weighted by $N(\hat{H}_i)$. To be more clearly describe the obtained weight for each face region, we show the weight of each face region in Fig.2. The brightness of the regions indicates their importance. That is, the more bright the region is, the higher weight it obtained (the more important). As can be seen, the brightest regions (corresponding to high weights) are in the area of two eyes, nose and mouth, while the darkest regions (corresponding to low weights) are around the cheek. This indicates that the learned weights can discover the more discriminative face regions to some extent and well capture the spatial information.

## 4 Discriminative Spatial Information based Sparse Coding

By imposing the weighted spatial information into the sparse coding scheme, our method can be formulated as follows:

$$\min_{\boldsymbol{\alpha}} \|\boldsymbol{W}(\boldsymbol{y} - \boldsymbol{D}\boldsymbol{\alpha})\|_2^2 + \lambda \sum_{j=1}^{C} \|\alpha_j\|_2 \tag{7}$$

where $C$ is the number of classes in the data and $\boldsymbol{W}$ is a diagonal matrix learned by Eq.6. The $i$-th diagonal coefficient in $\boldsymbol{W}$ is corresponding to $N(\hat{H}_i)$ of the $i$-th face image region. This weighting matrix will be used in sparse coding to incorporate discriminative abilities of different pixel locations. Eq.7 has the following advantage: some regions rich in texture, like areas around eyes, should contain more information; thus they will be adaptively assigned with high weights to ensure very small residuals. While others, like cheek, can be almost homogenous, can be assigned with low weights to reduce their effects on the regression estimation so that sensitiveness to these regions can be greatly reduced.

## 5 Optimization Algorithm

Problem Eq.7 is a convex formulation and we seek the global optimal solution. In this section, an efficient algorithm is derived to solve this problem. The detailed algorithm is given in Algorithm 1. More specifically, the dictionary $\boldsymbol{D}$ and test sample $\boldsymbol{y}$ are spatially weighted as:

$$\begin{aligned} \boldsymbol{y}^* &= \boldsymbol{W}\boldsymbol{y} \\ \boldsymbol{D}^* &= \boldsymbol{W}\boldsymbol{D} \end{aligned} \tag{8}$$

Thus, Eq.7 can be rewritten as:

$$\min_{\boldsymbol{\alpha}} \|\boldsymbol{y}^* - \sum_{j=1}^{C} D_j^* \alpha_j\|_2^2 + \lambda \sum_{j=1}^{C} \|\alpha_j\|_2 \tag{9}$$

Let $j = 1, \dots, C$ and $n_j$ is the number of samples in $j$-th class, the solution $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_C)$ must satisfy the Karush-Kuhn-Tucker (KKT) conditions, i.e., a necessary and sufficient condition for $\alpha$ to be a solution to Eq.9 is

$$-D_j^{*'}(y^* - D^*\alpha) + \frac{\lambda \alpha_j \sqrt{n_j}}{\|\alpha_j\|} = 0 \quad \forall \alpha_j \neq 0 \tag{10}$$

$$\| - D_j^{*'}(y^* - D^*\alpha)\| \leq \lambda \sqrt{n_j} \quad \forall \alpha_j = 0 \tag{11}$$

Recall that $D_j^{*'} D_j^* = I_{n_j}$. It can be easily verified that the solution to Eq.10 and Eq.11 is

$$\alpha_j = \left(1 - \frac{\lambda \sqrt{n_j}}{\|S_j\|}\right)_+ S_j, \tag{12}$$

where $S_j = D_j^{*'}(y - D^*\alpha_{-j})$, with $\alpha_{-j} = (\alpha_1^{'}, \dots, \alpha_{j-1}^{'}, 0^{'}, \alpha_{j+1}^{'}, \dots, \alpha_C^{'})$. The solution to Eq.9 can therefore be obtained by iteratively applying Eq.12 to $j = 1, \dots, C$.

The algorithm is found to be very stable and usually reaches a reasonable convergence tolerance within a few iterations. However, the computational burden increases dramatically as the number of predictors increases. Therefore, we used group least angle regression selection (LARS) to solve the problem and can be summarized as Algorithm 1.

**Algorithm 1:** Discriminative Spatial Information based Sparse Coding(DSISC)

```
Input: test image $y$, dictionary $D$
Output: sparse code $\alpha$
1: start from $\alpha^{[0]} = 0, k = 1$ and $r^{[0]} = y$;
2: compute the current ‘most correlated set’
        $B_1 = \arg\max_j \|D_j^{*'} r^{[k-1]}\|^2/n_j$        (13)
3: compute the current direction $\gamma$ which is a $n = \sum n_j$ dimensional vector with $\gamma_{B_k^c} = 0$ and
        $\gamma_{B_k} = (D_{B_k}^{*'} D_{B_k}^*)^{-1} D_{B_k}^{*'} r^{[k-1]},$        (14)
   where $D_{B_k}^*$ denotes the matrix comprised of the columns of $D^*$ corresponding to $B_k$.
4: for every $j \notin B_k$, compute how far the group LARS algorithm will progress in direction $\gamma$
   before $D_j^*$ enters the most correlated set. This can be measured by an $\beta_j \in [0,1]$ such that
        $\|D_j^{*'}(r^{[k-1]} - \beta_j D^*\gamma)\|^2/n_j = \|D_{j'}^{*'}(r^{[k-1]} - \beta_j D^*\gamma)\|^2/n_{j'}$        (15)
   where $j^{'}$ is arbitrarily chosen from $B_k$.
5: if $B_k \neq \{1, \ldots, C\}$, let $\beta = \min_{j \notin B_k} (\beta_j) \equiv \beta_{j^*}$ and update $B_{k+1} = B \bigcup \{j^*\}$; otherwise set
   $\beta = 1$.
6: update $\alpha^{[k]} = \alpha^{[k-1]} + \beta\gamma, \gamma^{[k]} = y - D\alpha^{[k]}$ and $k = k + 1$. Go back to Step 3 until
   $\beta = 1$.
```

## 6 Classification

Once Eq.9 is minimized, the resulting coefficient vector $\alpha$ is used as the feature vector for the test image $\boldsymbol{y}$. For classification, we first calculate our proposed DSISC of the query input, and we recognize this input as the class with the lowest reconstruction error using only the associated coefficient attributes in $\alpha_j$. The decision process is shown as follows:

$$j^\star = \arg\min_j \|\boldsymbol{y} - D_j^* \alpha_j\|_2^2 \tag{16}$$

where $D_j^* = [d_{j1}^*, d_{j2}^*, \ldots, d_{jn_j}^*] \in \Re^{m \times n_j}$ contains the training samples form the $j$-th class.

Suppose $\boldsymbol{W}$ is a diagonal matrix of weights assigned to pixels of the query image $\boldsymbol{y}$. More specifically, $\boldsymbol{W}$ can be represented by the weight of different face regions as follows,

$$\boldsymbol{W} = \boldsymbol{W_{eyes}} + \boldsymbol{W_{nose}} + \boldsymbol{W_{cheek}} + \boldsymbol{W_{other}}$$

where $\boldsymbol{W_{eye}}$, $\boldsymbol{W_{nose}}$, $\boldsymbol{W_{cheek}}$ and $\boldsymbol{W_{other}}$ are diagonal matrix of weights corresponding to face regions of eye, nose, cheek and other part respectively. For classification, we first spatially weighted the test image $\boldsymbol{y}$ as

$$\boldsymbol{y^*} = \boldsymbol{Wy} = \boldsymbol{W_{eyes}y} + \boldsymbol{W_{nose}y} + \boldsymbol{W_{cheek}y} + \boldsymbol{W_{other}y}$$

Then, we recognize this input test image as the class with the lowest reconstruction error or coding residual (see Eq.16). The decision process is shown as follows:

$$j^{\star} = \arg\min_{j} \|\boldsymbol{y^{*}} - D_j^{*}\alpha_j\|_2^2$$

For example, there are two classes ($j = 1$ or $j = 2$) in the data and the test image $\boldsymbol{y}$ is belong to the $j = 1$ class. For classification, we compare the following two equations

$$\begin{aligned}
j_1^{\star} = \|\boldsymbol{y^{*}} - D_1^{*}\alpha_1\|_2^2 &= \|(\boldsymbol{y} - D_1\alpha_1)\boldsymbol{W_{eyes}}\|_2^2 + \|(\boldsymbol{y} - D_1\alpha_1)\boldsymbol{W_{nose}}\|_2^2 \\
&+ \|(\boldsymbol{y} - D_1\alpha_1)\boldsymbol{W_{cheek}}\|_2^2 + \|(\boldsymbol{y} - D_1\alpha_1)\boldsymbol{W_{other}}\|_2^2 \\
j_2^{\star} = \|\boldsymbol{y^{*}} - D_2^{*}\alpha_2\|_2^2 &= \|(\boldsymbol{y} - D_2\alpha_2)\boldsymbol{W_{eyes}}\|_2^2 + \|(\boldsymbol{y} - D_2\alpha_2)\boldsymbol{W_{nose}}\|_2^2 \\
&+ \|(\boldsymbol{y} - D_2\alpha_2)\boldsymbol{W_{cheek}}\|_2^2 + \|(\boldsymbol{y} - D_2\alpha_2)\boldsymbol{W_{other}}\|_2^2
\end{aligned}$$

Assume that we obtain the coding residual (or reconstruction error) for right class is $j_1^{\star} = 2$ and coding residual for wrong class is $j_2^{\star} = 4$, so we recognize the test image as class 1 (j=1). We also expect that the difference between coding residual for right class and coding residual for wrong class is as large as possible, which indicating more discriminative power.

## 7 Experiments and Comparisons

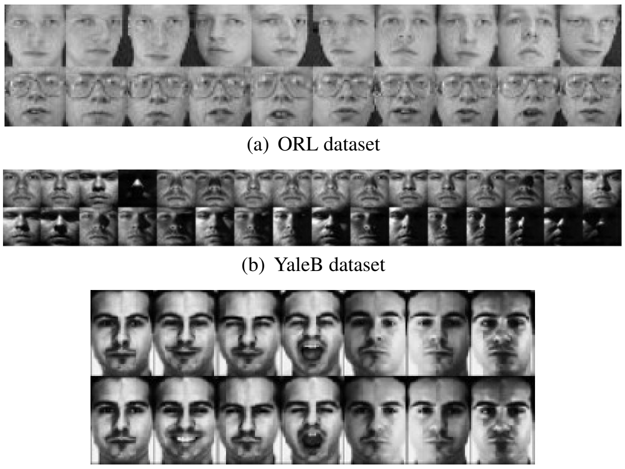
To demonstrate the effectiveness and robustness of the proposed approach DSISC, we conduct experiments on three benchmark face data sets, i.e., ORL [16], the Extended YaleB [7] and AR [12]. Table. 1 summarizes the extents and properties of the three face data-sets. All the face images are cropped and aligned by using the locations of eyes which are provided by the face databases. For all methods, the training samples are used as the dictionary $\boldsymbol{D}$ in sparse coding. In Fig.3, we show the closely cropped images and these all contain facial structure.

**Table 1** Summary of three benchmark face data sets

| Data-set | Sample | Features | Classes |
|---|---|---|---|
| ORL | 400 | 112*92 | 40 |
| YaleB | 2414 | 54*48 | 38 |
| AR | 1400 | 60*43 | 100 |

**ORL dataset:** it contains 40 distinct individuals with ten images per person. The images are taken at different time instances, and include variations in facial expression and facial detail (glasses/no glasses), as shown in Fig.3(a). The size of each cropped image is 112×92. For each subject, we select $t = 5$ images for training and use the rest for test.

**YaleB dataset:** The Extended YaleB database contains 16128 images of 38 human subjects under 9 poses and 64 illumination conditions. In this experiment, we choose the frontal pose and use all the images under different illumination. Finally,



(a) ORL dataset

(b) YaleB dataset

(c) AR dataset

**Fig. 3** The sample of cropped face images from three face datasets.

we get 2414 frontal face images of 38 individuals in total. All the images are manually aligned and cropped. The size of each cropped image is $54 \times 48$, which were taken under varying illumination conditions. We randomly split the database into two halves. One half (about 32 images per person) was used as the dictionary, and the other half for testing.

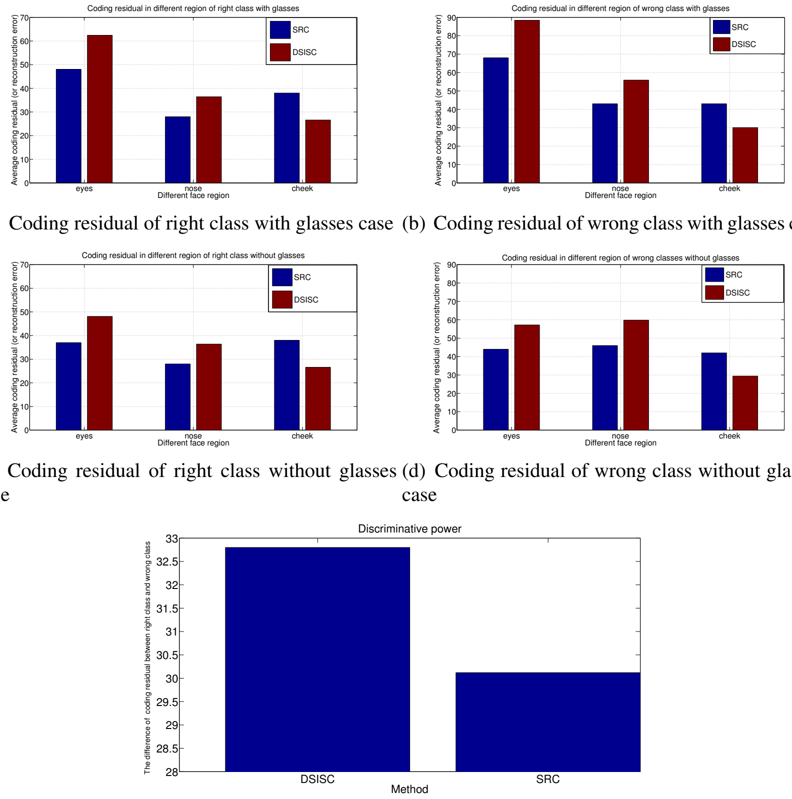
**AR dataset:** it consists of 4000 frontal images of 126 subjects. In this experiments, a subset (with only illumination and expression changes) that contains 50 male subjects and 50 female subjects is chosen from the AR database. As in [19], for each subject, we choose $t = 7$ images for training and take the rest for test. The images are cropped to 60×43 pixels (see Fig.3(c)).

In order to explore the discriminative capabilities of the information captured by our method, we compare the classification results from our proposed method (DSISC) with five representative algorithms like nearest neighbor (NN), nearest subspace (NS), linear support vector machine (SVM), the recently developed sparse coding methods SRC [19] and RSC [21]. In the experiments, PCA is used to reduce the dimensionality of original face features., and the Eigenface features are used for all the competing methods.

## 7.1 The contribution of spatial weighting strategy

The aim of this experiment is to demonstrate the discriminative ability of placing weight values on the regression estimation.

To demonstrate the effectiveness and robustness of the proposed approach DSISC, we conduct experiments on AR dataset. For testing, we randomly choose 10 images with glasses and 10 image without glasses. The final coding residual (or reconstruction error) is computed by averaging of all the test images. In order to explore the discriminative capabilities of the information captured by adding weight values in our method, we compare the classification results from our proposed method (DSISC) with alternative sparse coding method SRC [19] without adding any weights.



(a) Coding residual of right class with glasses case (b) Coding residual of wrong class with glasses case

(c) Coding residual of right class without glasses case (d) Coding residual of wrong class without glasses case

(e) The difference of coding residual between right class and wrong class

**Fig. 4** The performance of discriminative ability in DSISC and SRC.

Coding residual (or reconstruction error) of DSISC and SRC are shown in Fig.4. From Fig.4 (a), (b), (c) and (d), we can observed that compared with SRC, the weight values used in our method DSISC increase the coding residual (or reconstruction error) in regions of eyes and nose. However, the coding residual (or reconstruction error) are increasing in both right class and wrong class (See Fig.4 (a) and (b) with glasses case, Fig.4 (c) and (d) without glasses case). The final classification accuracy depends on whether coding residual (or reconstruction error) of the right class is smaller than the coding residual (or reconstruction error) of the wrong class. Fig.4(d) shows the difference of coding residual (or reconstruction error) between right class and wrong class in two methods. It is clear that DSISC has larger difference of coding residual between right class and wrong class, which indicates more discriminative power

### 7.2 Face recognition without Occlusion

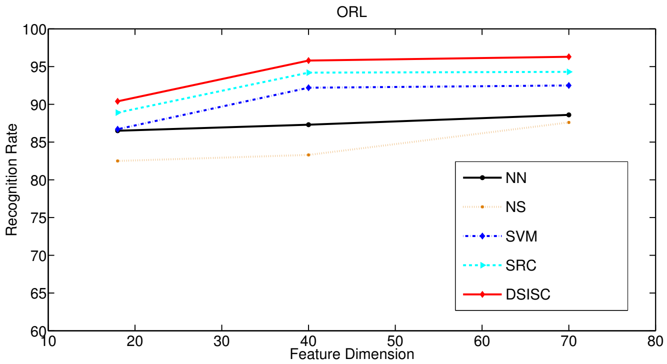


**Fig. 5** Face Recognition rates by the competing methods under different feature dimensions in ORL without occlusion

**Table 2** Face recognition rates on the ORL database

| Dim | 18 | 40 | 70 |
|---|---|---|---|
| NN | 86,5 % | 87.3 % | 88.6 % |
| NS | 82.5 % | 83.3% | 87.6% |
| SVM | 86.7% | 92.2% | 92.5% |
| SRC | 88.9% | 94.2% | 94.3% |
| DSISC | **90.4%** | **95.8%** | **96.3%** |

**Results on ORL databse**: Fig.5 compares the recognition rates of the competing methods under various feature dimensions. As Fig.5 shows, our method outperforms other baseline methods in all cases, and SRC performs the second best which verifies the effectiveness of sparse representation. SRC shows inferior performance to our DSISC. This indicates the advantage of jointly considering the spatial information and sparsity. For clear comparison, we summarize the recognition rates versus feature dimensions by different methods in Table 2. It is clear that our proposed method DSISC is, by and large, superior to the alternative methods in all dimensions. SRC performs the second best. The highest recognition rate of DSISC on 70 dimensions is 96.3%, more than 2% improvement over SRC.

**Results on YaleB databse**: From Fig.3 and Table 3, we can see that DSISC still maintain the best recognition rates at all levels. When the dimension is too low, NN and NS methods achieve very low recognition rate. SVM obtains much better results

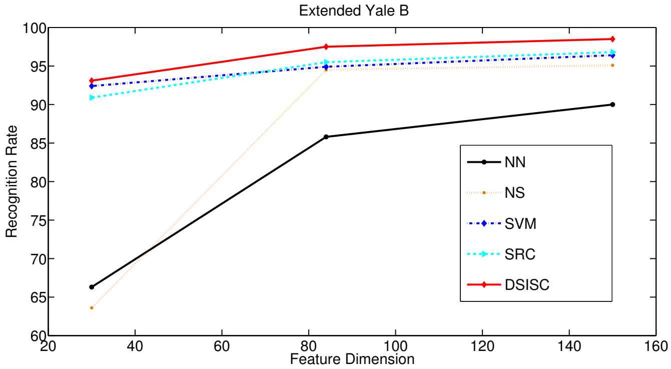


**Fig. 6** Face Recognition rates by the competing methods under different feature dimensions in YaleB without occlusion

**Table 3** Face recognition rates on the Extended YaleB database

| Dim | 30 | 84 | 150 |
|---|---|---|---|
| NN | 66.3 % | 85.8 % | 90 % |
| NS | 63.6 % | 94.5% | 95.1% |
| SVM | 92.4% | 94.9% | 96.4% |
| SRC | 90.9% | 95.5% | 96.8% |
| DSISC | **93.1%** | **97.5%** | **98.5%** |

compared with NN and NS, since there are relatively enough (32 per class) training samples. The recognition rates of DSISC and SRC are both at least 20% higher than NN and NS. This shows that sparse representation does have much contribution to face recognition.

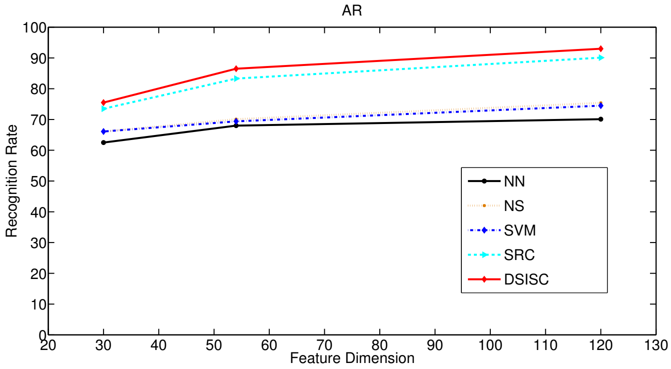


**Fig. 7** Face Recognition rates by the competing methods under different feature dimensions in AR without occlusion

**Table 4** Face recognition rates on the AR database

| Dim | 30 | 54 | 120 |
|---|---|---|---|
| NN | 62.5 % | 68 % | 70.1 % |
| NS | 66.1 % | 70.1% | 75.4% |
| SVM | 66.1% | 69.4% | 74.5% |
| SRC | 73.5% | 83.3% | 90.1% |
| DSISC | **75.5%** | **86.5%** | **93%** |

**Results on AR databse**: The recognition rates versus feature dimension by NN, NS, SVM, SRC and DSISC are shown in Fig.7. Again, we observe that our proposed method DSISC is always on the top. Table 4 lists the recognition rates by the competing methods. The results validate that DSISC ans SRC are the best in accuracy, with at least 15% improvement than the other three methods when the dimensionality is 120, but SRC is still inferior to our method. Nevertheless, when the dimension is too low, all the methods cannot achieve very high recognition rate. On other dimensions, DSISC outperforms SRC by about 3%. SVM does not give results in this experiment because there are not enough training samples (7 samples per class here) and there are high variations between training set and testing set. The maximal recognition rates of DSISC, SRC, SVM, NS and NN are 93%, 90.1%, 74.5%,75.4% and 70.1% respectively.

## 7.3 Face recognition with occlusion

In this subsection, we run extensive tests to verify the robustness and effectiveness of our method to different kinds of occlusions including random pixel corruption, random block occlusion, and real disguise.

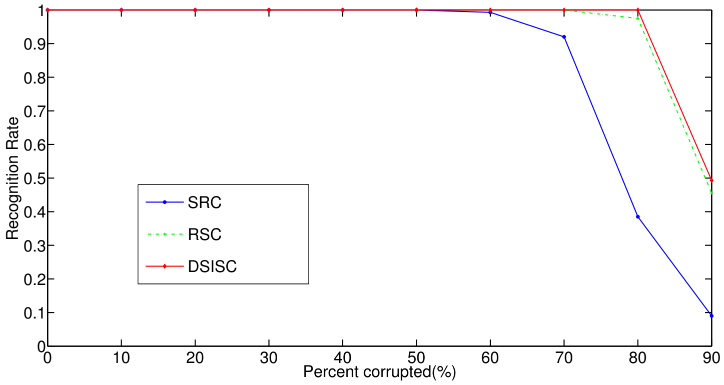


**Fig. 8** The recognition accuracies of our approach, the SRC, and the RSC under various percentages of random pixel corruption

**Table 5** The recognition accuracies of our approach, the SRC, and the RSC under various percentages of random pixel corruption

| Corrupted | 0 | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% |
|---|---|---|---|---|---|---|---|---|---|---|
| SRC | 1 | 1 | 1 | 1 | 1 | 1 | 0.993 | 0.92 | 0.385 | 0.09 |
| RSC | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.975 | 0.455 |
| DSISC | **1** | **1** | **1** | **1** | **1** | **1** | **1** | **1** | **1** | **0.463** |

**Face recogntion with pixel corruption**: The Extended YaleB database is used for this purpose. In accordance to the experiments in [19, 21], we use Subset 1 and 2 (717 images, normal-to-moderate lighting conditions from) of the Extended YaleB database for training, and used Subset 3 (453 images with more extreme lighting conditions) for testing. All the images were resized to 96×84. For each testing image, we used random values within [0,255] to replace a certain percentage of pixels in the image to simulate pixel corruption. Locations of the corrupted pixels are random and unknown to the algorithm.

Fig.8 plots the recognition rates of the SRC, the RSC, and our method under various percentages (0% - 90%) of corrupted pixels. For clear comparison, we summarize the recognition rates of different methods in Table 5. We can see that all three methods reports an almost perfect accuracy when the percentage of corruption is between (0% - 60%). When 80% of pixels are corrupted, our method can still classify all the test images correctly, while the RSC has a recognition rate of 97.5%, and the SRC only has a recognition rate of 38.5%. The recognition rates of the SRC, the RSC and our method are, respectively, 9%, 45.5%, 47.3% when 90% of the pixels are corrupted.

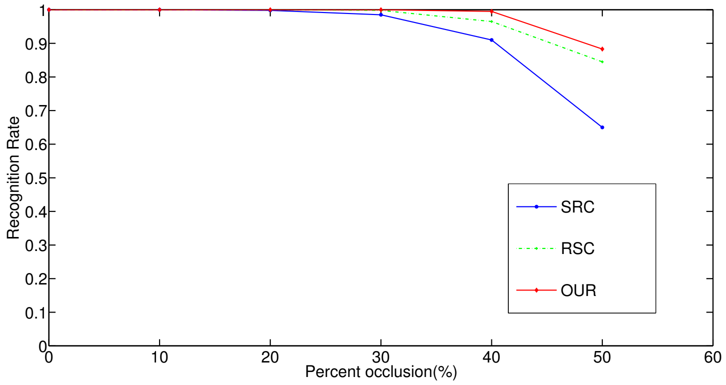


**Fig. 9** The recognition accuracies of our approach, the SRC, and the RSC under different levels of block occlusion

**Face recogntion with block occlusion**: Fig.9 plots the recognition rates of the SRC, the RSC, and our method under various percentages (0% - 50%) of block occlusion. To test the robustness of our method to artificial block occlusion, we randomly chose a square block in each test image and replaced it with an irrelevant image. As in

**Table 6** The recognition accuracies of our approach, the SRC, and the RSC under different levels of block occlusion

| Occlusion | 0 | 10% | 20% | 30% | 40% | 50% |
|---|---|---|---|---|---|---|
| SRC | 1 | 1 | 0.998 | 0.985 | 0.91 | 0.65 |
| RSC | 1 | 1 | 1 | 0.998 | 0.965 | 0.845 |
| DSISC | **1** | **1** | **1** | **1** | **0.995** | **0.883** |

[19, 21], Subset 1 and Subset 2 of Extended Yale B were used for training and Subset 3 for testing. All the images were cropped to 96×84. The results of the SRC, the RSC and our method are shown in Table 6. Again we see that when the block occlusion is 40%, the recognition rate of our method is very close to 100%. When half of the image is occluded, our method achieves a recognition rate of 88.3%, over 3% higher than that of the RSC, while the SRC only achieves a rate of 65%.

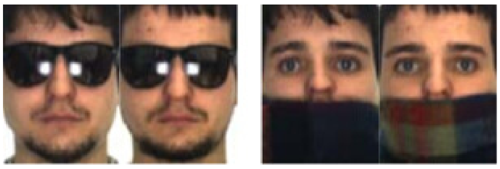
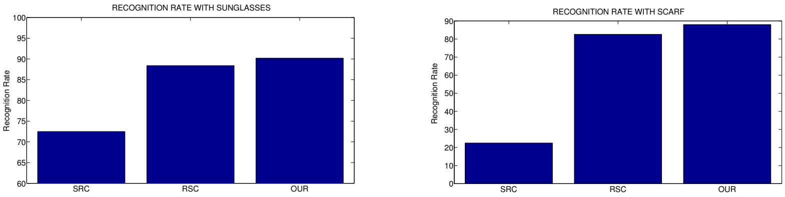


**Fig. 10** The testing samples with sunglasses and scarves in the AR database



(a) Sunglasses

(b) Scarf

**Fig. 11** Face recognition rates of the SRC, the RSC and our method on the AR database with occlusion of real disguise.

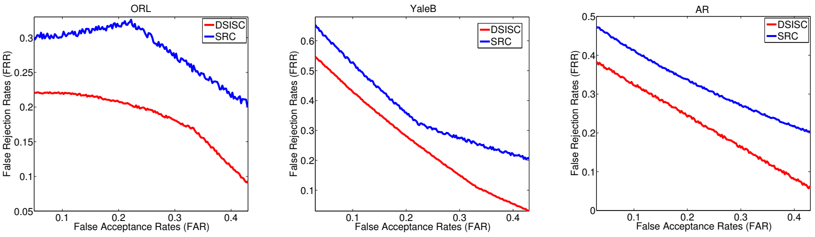
**Table 7** Face recognition rates of the SRC, the RSC and our method on the AR database with occlusion of real disguise

| Algorithm | SRC | RSC | DSISC |
|---|---|---|---|
| Sunglasses | 72.5% ± 8.6 | 88.4% ± 6.6 | **90.2% ± 4.3** |
| Scarf | 22.5% ± 7.3 | 82.6% ± 4.8 | **88% ± 3.6** |

**Face recogntion with real face disguise**: We test our proposed method DSISC's ability to cope with real possibly malicious occlusions using a subset of the AR face database. As in [19], the subset consists of 1400 images from 100 subjects, 50 male and 50 female. For training, we choose 400 images (about 4 samples per subject) of non-occluded frontal views with various facial expression. For testing, we randomly choose 300 images with sunglasses and scarves (as shown in Fig. 10), and this procedure is repeated 10 times. The final accuracy is computed by averaging of the accuracies from all experiments.

The results are shown in Fig.11 and Table 7. It can be seen that even with sunglasses or scarfs, our method still maintain the best recognition rates in all situations. Moreover, for the case of face recognition with scarf, DSISC significantly outperforms SRC by a margin of 66%. Compared to RSC, our method still gets competing results, 2% higher in face recognition with sunglasses, while in face recognition with scarf, much more improvement is obtained (6% higher than that of RSC). The results further verify that robust face recognition can be obtained by considering the discriminative spatial information of face images in sparse coding.

### 7.4 Face Recognition System Evaluation



(a) ORL (b) YaleB (c) AR

**Fig. 12** FAR/FRR plot of face recognition on three databases.

In the presented experiment the ORL database [16], the Extended YaleB [7] and AR [12] databases have been used. The experiments were conducted using an i5 processor with MATLAB R2013b. The recognition speed of proposed method DSISC on three datasets are 0.36 s (ORL), 1.12 s (YaleB) and 0.57 s (AR) respectively.

The first database (ORL database) includes 40 distinct individuals with 10 images per person. We randomly select 20 individuals (about 200 images) as reference images, and the other 20 individuals for testing. The second database (YaleB database) contains 2414 frontal face images of 38 individuals in total. We randomly select 19 individuals (about 1207 images) as reference images, and the other 19 individuals for testing. In the case of the AR database, we randomly select 50 individuals (about 700 images) as reference image, and the other 50 individuals for testing. The split procedure for each database is repeated 10 times. The final false acceptance error rate (FAR) and the false rejection error rate (FRR) are computed by averaging the corresponding error rates from each of the random subsets.

From the DET curves in Fig.12, which plot the false acceptance error rate against the false rejection error rate at various values of the decision threshold, we can find that the proposed method DSISC significantly outperforms SRC on three datasets.

**Table 8** Performance of proposed method DSISC (with face image dimension variation)

| Dim | 30 | | 40 | | 50 | |
|---|---|---|---|---|---|---|
| | recall | precision | recall | precision | recall | precision |
| ORL | 90.4% | 89.6% | 91.1% | 89.2% | 87.6% | 91.8% |
| YaleB | 92.4% | 91.6% | 92.4% | 90.5% | 88.2% | 89.1% |
| AR | 90.6% | 91.4% | 90.2% | 92.1% | 85.5% | 88.6% |

**Table 9** Performance of SRC (with face image dimension variation)

| Dim | 30 | | 40 | | 50 | |
|---|---|---|---|---|---|---|
| | recall | precision | recall | precision | recall | precision |
| ORL | 81.3% | 89.2% | 90.2% | 84.1% | 78.8% | 92.7% |
| YaleB | 86.3% | 90.9% | 89.9% | 82.2% | 83.2% | 78.6% |
| AR | 86.4% | 81.2% | 78.1% | 87.1% | 91.1% | 87.9% |

In order to fully evaluate the proposed method for face recognition, the precision and recall are calculated. Table 8 and Table 9 show the average precision and recall rate by varying the dimension of face images. It is clear that our proposed method DSISC shows superior performance over SRC in all cases.

## 8 Conclusion

In this paper, we proposed to incorporate a discriminative spatial information into the sparse coding for face recognition. We learn the weights at face locations according to the information entropy in each face region, so as to highlight locations in face images that are important for classification. Furthermore, we consider the group structure of training images (i.e. those from the same subject) and added an $\ell_{2,1}$-norm (group Lasso) constraint upon the formulation, which enforcing the sparsity at the group level. Finally, an efficient group Least Angle Regression Selection (LARS) is presented to solve the resulting group sparse optimization problem.

## Acknowledgment

This work is supported by National Natural Science Foundation of China (Grant No.61402389).

## References

1. Ahonen, T., Hadid, A., Pietikäinen, M.: Face recognition with local binary patterns. In: ECCV 2004, pp. 469–481. Springer (2004)
2. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.: Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence **19**(7), 711–720 (1997)
3. Chai, Z., Mendez-Vazquez, H., He, R., Sun, Z., Tan, T.: Semantic pixel sets based local binary patterns for face recognition. In: ACCV, pp. 639–651. Springer (2013)
4. Cover, T.M., Thomas, J.A.: Elements of information theory. John Wiley & Sons (2012)
5. Gao, S., Tsang, I.W., Chia, L.T., Zhao, P.: Local features are not lonely–laplacian sparse coding for image classification. In: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3555–3561. IEEE (2010)
6. Gao, S., Tsang, I.W.H., Chia, L.T.: Kernel sparse representation for image classification and face recognition. In: ECCV 2010, pp. 1–14. Springer (2010)
7. Georghiades, A.S., Belhumeur, P.N., Kriegman, D.: From few to many: Illumination cone models for face recognition under variable lighting and pose. IEEE Transactions on Pattern Analysis and Machine Intelligence **23**(6), 643–660 (2001)
8. He, X., Niyogi, P.: Locality preserving projections. In: Advances in Neural Information Processing Systems, pp. 153–160 (2004)
9. He, X., Yan, S., Hu, Y., Niyogi, P., Zhang, H.J.: Face recognition using laplacianfaces. IEEE Transactions on Pattern Analysis and Machine Intelligence **27**(3), 328–340 (2005)
10. Ji, S., Xue, Y., Carin, L.: Bayesian compressive sensing. IEEE Transactions on Signal Processing **56**(6), 2346–2356 (2008)
11. Liu, Y., Wu, F., Zhang, Z., Zhuang, Y., Yan, S.: Sparse representation using nonnegative curds and whey. In: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3578–3585. IEEE (2010)
12. Martinez, A.M.: The ar face database. CVC Technical Report **24** (1998)
13. Peng, Y., Ganesh, A., Wright, J., Xu, W., Ma, Y.: Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images. IEEE Transactions on Pattern Analysis and Machine Intelligence **34**(11), 2233–2246 (2012)
14. Quanz, B., Huan, J., Mishra, M.: Knowledge transfer with low-quality data: A feature extraction issue. IEEE Transactions on Knowledge and Data Engineering **24**(10), 1789–1802 (2012)
15. Ramirez, I., Sapiro, G.: Universal sparse modeling. Tech. rep., DTIC Document (2010)
16. Samaria, F.S., Harter, A.C.: Parameterisation of a stochastic model for human face identification. In: Proceedings of the Second IEEE Workshop on Applications of Computer Vision, 1994., pp. 138–142. IEEE (1994)
17. Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., Gong, Y.: Locality-constrained linear coding for image classification. In: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3360–3367. IEEE (2010)

18. Wright, J., Ma, Y.: Dense error correction via $\ell_1$-minimization. IEEE Transactions on Information Theory **56**(7), 3540–3560 (2010)
19. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust face recognition via sparse representation. IEEE Transactions on Pattern Analysis and Machine Intelligence **31**(2), 210–227 (2009)
20. Yang, J., Yu, K., Gong, Y., Huang, T.: Linear spatial pyramid matching using sparse coding for image classification. In: IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 1794–1801. IEEE (2009)
21. Yang, M., Zhang, D., Yang, J.: Robust sparse coding for face recognition. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 625–632. IEEE (2011)
22. Yang, M., Zhang, L., Yang, J., Zhang, D.: Regularized robust coding for face recognition. IEEE Transactions on Image Processing **22**(5), 1753–1766 (2013)
23. Zhang, C., Wang, S., Huang, Q., Liu, J., Liang, C., Tian, Q.: Image classification using spatial pyramid robust sparse coding. Pattern Recognition Letters **34**(9), 1046–1052 (2013)
24. Zhang, D., Yang, M., Feng, X.: Sparse representation or collaborative representation: Which helps face recognition? In: 2011 IEEE International Conference on Computer Vision, pp. 471–478. IEEE (2011)
25. Zheng, M., Bu, J., Chen, C., Wang, C., Zhang, L., Qiu, G., Cai, D.: Graph regularized sparse coding for image representation. IEEE Transactions on Image Processing **20**(5), 1327–1336 (2011)

Zhihong Zhang received his BSc degree (1st class Hons.) in computer science from the University of Ulster, UK, in 2009 and the PhD degree in computer science from the University of York, UK, in 2013. He won the K. M. Stott prize for best thesis from the University of York in 2013. He has published 18 papers in international journals and conferences. He is now an assistant professor at the software school of Xiamen University, China. His research interests are wide-reaching but mainly involve the areas of pattern recognition and machine learning, particularly problems involving graphs and networks.

Lu Bai received the Ph.D. degree from the University of York, York, UK, and both the B.Sc. and M.Sc degrees from Faculty of Information Technology, Macau University of Science and Technology, Macau SAR, P.R. China. He is now a Asistant Professor in School of Information, Central University of Finance and Economis, Beijing, China. His current research interests include structural pattern recognition, machine learning, quantum walks on networks and graph matching, especially in kernel methods and complexity analysis on (hyper)graphs and networks.

Yuanheng Liang received the B.S degree from South China University of technology in 2013.He is currently pursuing the M.S degree in Xiamen University His currently research interests include pattern recognition and machine learning.

Edwin R. Hancock received the B.Sc., Ph.D., and D.Sc. degrees from the University of Durham, Durham, U.K. He is now a Professor of computer vision in the Department of Computer Science, University of York, York, U.K. He has published nearly 150 journal articles and 550 conference papers. Prof. Hancock was the recipient of a Royal Society Wolfson Research Merit Award in 2009. He has been a member of the editorial board of the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, PATTERN RECOGNITION, COMPUTER VISION AND IMAGE UNDERSTANDING, and IMAGE AND VISION COMPUTING. His awards include the Pattern Recognition Society Medal in 1991, outstanding paper awards from the Pattern Recognition Journal in 1997, and the best conference best paper awards from the Computer Analysis of Images and Patterns Conference in 2001, the Asian Conference on Computer Vision in 2002, the International Conference on Pattern Recognition (ICPR) in 2006, British Machine Vision Conference (BMVC) in 2007, and the International Conference on Image Analysis and Processing in 2009. He is a Fellow of the International Association for Pattern Recognition, the Institute of Physics, the Institute of Engineering and Technology, and the British Computer Society. He was appointed as the founding Editor-in-Chief of the Institute of Engineering & Technology Computer Vision Journal in 2006. He was a General Chair for BMVC in 1994 and the Statistical, Syntactical and Structural Pattern Recognition in 2010, Track Chair for ICPR in 2004, and Area Chair at the European Conference on Computer Vision in 2006 and the Computer Vision and Pattern Recognition in 2008. He established the energy minimization methods in the Computer Vision and Pattern Recognition Workshop Series in 1997.

Author Photographs